CONTENTS

4 Model Selection

4.1 Method of Maximum Likelihood

Definition 1. Likelihood Function The joint density function of n random variables $X_1, \ldots, , X_n$ evaluated at $x_1, \ldots, x_n$, say $f(x_1, \ldots, x_n; \theta)$, is referred to as a likelihood function. For fixed $x_1, \ldots, x_n$ the likelihood function is a function of θ and often is denoted by $L(\theta)$. If $X_1, \ldots, , X_n$ represents a random sample from $f(x_1, \ldots, x_n; \theta)$, then

$$L(\theta) = f(x_1; \theta) \cdots f(x_n; \theta)$$

Definition 2. Maximum Likelihood Estimator Let $L(\theta) = f(x_1, \ldots, x_n; \theta)$, be the joint pdf of $X_1, \ldots, , X_n$. For a given set of observations, $(x_1, \ldots, x_n)$, a value $\hat{\theta}$ in Ω at which $L(\theta)$ is a maximum is called a **maximum likehhood estimate** (MLE) of θ. That is $\hat{\theta}$ is a value of θ that satisfies

$$f(x_1, \ldots, x_n, ; \hat{\theta}) = \max_{\theta \in \Omega} f(x_1, \ldots, x_n; \theta).$$

Note:

1. If each set of observations $(x_1, \dots, x_n)$ corresponds to a unique value $\hat{\theta}$, then this procedure defines a function, $\hat{\theta} = t(x_1, \dots, x_n)$. This same function when applied to the random sample, $\hat{\theta} = t(X_1, \dots, X_n)$ is called the **maximum likelihood estimator**, also denoted MLE.

2. Any value of $\hat{\theta}$ that maximizes $L(\theta)$ also will maximize the log-likelihood, $lnL(\theta) = l(\theta)$, so for computational convenience then alternate form of the maximum likelihood equation,

$$\frac{d}{d\theta} l(\theta)$$

often will be used.

Example 1.
Find the MLEs based on random sample $X_1, \dots, X_n$ from each of the following distributions:

(a) $X_i \sim POI(\lambda)$

(b) $X_i \sim EXP(\theta)$

(c) $X_i \sim N(\mu, \sigma^2)$

(d) $X_i \sim Pareto(\alpha, \theta = 100)$

(e) $X_i \sim U(0, \theta)$

Example 2.
You are given:

$$f(x) = \frac{1}{\theta}e^{-(\frac{x-\eta}{\theta})}, x \geq \eta$$

for $\theta > 0$ and $\eta \in \mathbf{R}$. Suppose that $X_1, X_2, \ldots, X_n$ are iid with pdf $f(x|\theta, \eta)$. Determine the maximum likelihood estimators of η and θ.

4.2 Goodness of Fit Tests

The goodness of fit (GOF) tests measures the compatibility of a random sample with a theoretical probability distribution function. In other words, these tests show how well the distribution you selected fits to your data.

4.2.1 Chi-Square Goodness of Fit Test

- **Known Parameter Case**

  To test $H_0 : X \sim F(x)$.
  Group the data if not already grouped. For each group, say $A_1, \ldots, A_k$, let $p_j = P(X \in A_j)$ where $X \sim F(x)$. Let n_j be the number of observations in group j, so $n = \sum_{j=1}^{k} n_j$ and under H_0, the expected number in the j^{th} group is $E = np_j$.

  The chi-square statistic is

  $$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - E_j)^2}{E_j}$$

  $H_0 : X \sim F(x)$ is rejected if

  $$\chi^2 \geq \chi^2_{(1-\alpha)}(k-1).$$

  As a general principle, as many groups as possible should be used to increase the number of degrees of freedom, as long as $E_j \geq 5$.

- **Unknown Parameter Case**

Suppose we wish to test $H_0 : X \sim f(x; \theta_1, \cdots, \theta_p)$ where there are p unknown parameters. To compute the χ^2 statistic, the expected number under H_0 now must be estimated. Then the unknown $p_j = P(X \in A_j)$ are functions of $\theta_1, \cdots, \theta_p$. If MLE is used to estimate $\theta_1, \cdots, \theta_p$. then the limiting distribution of the χ^2 statistic is chi-squares with degrees of freedom $k-1-p$. That is, approximately

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - E_j)^2}{E_j} \sim \chi^2(1-\alpha)(k-1-p),$$

where $Ej = n\hat{p}_j$.

Example 3 (T4Q1)**.**

You are given the following observed claim frequency data collected over a period of 365 days:

Number of Claims per Day	Observed Number of Days
0	51
1	130
2	117
3	67
4+	0

Fit a Geometric distribution to the above data, using the method of maximum likelihood. Group the data by number of claims per day into four groups:

0 1 2 3 or more

Apply the chi square goodness of fit test to evaluate the null hypothesis that the claims follow a Geometric distribution. Let Q be the value of the chi-square statistic and u be the degrees of freedom. Determine $Q - u$.

Example 4 (T4Q2).

You are given the following:

- 1090 observed losses have been recorded an are grouped as follows:

Interval	Number of Losses
[0,1)	150
[1,5)	340
[5,10)	230
[10,15)	180
[15, ∞)	190

- The random variable X underlying the observed losses, is believed to follow the gamma disribution with $\alpha = 2$ and $\theta = 5$.

Determine the value of Pearson's goodness-of-fit statistic.

Example 5 (T4Q3).

You are given the following claim frequency data:

Number of Claims	0	1	2	3	4
Number of risks	9	12	12	18	5

The null hypothesis is that the number of claims per risk follows a uniform distribution on 0, 1, 2, 3, and 4. Let Q be the value of the chi-square statistic and u be the degrees of freedom. Determine $Q + u$.

4.2.2 Komogrov-Smirnov Test

Let t be the left truncation point ($t = 0$ if there is no truncation) and let u be the right censoring point ($u = \infty$ if there is no censoring). Then, the test statistic is

$$D = max|F_n(x) - F^*(x)|.$$

This test should only be used on individual data. This is to ensure that the step function $F_n(x)$ is well defined. Also, the model distribution function $F^*(x)$ is assumed to be continuous over the relevant range.

Commonly used critical values for this test are

α	0.10	0.05	0.025	0.01
Critical Value	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.48}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Example 6 (T4Q4)**.**

A random sample of 5 claims $x_1, \ldots, x_5$ is taken from the probability density function

$$f(x_i) = \frac{\alpha\theta^\alpha}{(x_i + \theta)^{\alpha+1}}, \alpha, \theta, x_i > 0.$$

In ascending order the observations are: 66, 154,246, 490, 883
Suppose the parameters are $\alpha = 3$ and $\theta = 740$. Commonly used critical values for this test are

α	0.10	0.05	0.025	0.01
Critical Value	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.48}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Determine the result of the test at 0.1 significant level.

Example 7 (T4Q5).

A random sample of 10 claims $x_1, \dots, x_{10}$ is taken from the probability density function

$$f(x_i) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x_i^{\alpha-1} e^{-\frac{x_i}{\theta}}, x_i > 0.$$

In ascending order the observations are: 19.38, 45.06, 48.06, 56.1, 60.01, 88.95, 119.79, 123.32, 129.24, 277.14

Suppose the parameters are $\alpha = 4$ and $\theta = 51$. Commonly used critical values for this test are

α	0.10	0.05	0.025	0.01
Critical Value	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.48}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

Determine the result of the test at 0.1 significant level.

Example 8 (T4Q6).

You observe the following seven losses on a coverage with deductible 600 and maximum covered loss 12000:

$$657 \quad 887 \quad 1,344 \quad 1,709 \quad 2,257 \quad 2,741 \quad 4,270$$

In addition, you observe two losses above 12,000 for which the payments of 11,400 were made. You fit these losses to a single parameter Pareto distribution with $\theta = 3,000$ and $\alpha = 4$. Calculate the Kolmogorov-Smirnov statistic for the hypothesis.

Example 9 (T4Q7).

You observe the following seven losses on a coverage with deductible 600 and maximum covered loss 11000:

$$652 \quad 865 \quad 1,544 \quad 1,824 \quad 2,332 \quad 2,876 \quad 4,446$$

In addition, you observe two losses above 11,000 for which the payments of 10,400 were made. You fit these losses to a Weibull distribution with $\theta = 2,000$ and $\tau = 3$. Calculate the Kolmogorov-Smirnov statistic for the hypothesis.

4.2.3 Likelihood Ratio Test

The likelihood-ratio test assesses the goodness of fit of two competing statistical models based on the ratio of their likelihoods, specifically one found by maximization over the entire parameter space and another found after imposing some constraint. If the constraint (i.e., the null hypothesis) is supported by the observed data, the two likelihoods should not differ by more than sampling error. Thus the likelihood-ratio test tests whether this ratio is significantly different from one, or equivalently whether its natural logarithm is significantly different from zero.
The likelihood ratio test statistic for testing

$$H_0 : \theta \in \Theta_0 \text{ vs } H_1 : \theta \notin \Theta_0$$

is given by:

$$LR = -2\ln\left[\frac{\max_{\theta\in\Theta_0} L(\theta)}{\max_{\theta\in\Theta} L(\Theta)}\right]$$

where the quantity inside the brackets is called the likelihood ratio.
Often the likelihood-ratio test statistic is expressed as a difference between the log-likelihoods,

$$LR = -2[l(\theta_0) - l(\hat{\theta})],$$

where $\theta_0 \in \Theta_0$ and $\hat{\theta} \in \Theta$
Notes:

- A free parameter is one that is not specified, and that is therefore maximized using maximum likelihood.
- The number of degrees of freedom for the likeihood ratio test is the number of free parameters(r) in the alternative model, the model of alternative hypothesis, minus the number of free parameters in the base model, the model of null hypothesis.
- The reason for multiplying by negative two is mathematical so that, by Wilks' theorem, LR has an asymptotic χ^2-distribution under the null hypothesis. Thus, an approximate size α test is to reject H_0 if $-[l(\theta_0) - l(\hat{\theta})] \geq \chi^2_{1-\alpha}(r)$.
- As all likelihoods are positive, and as the constrained maximum cannot exceed the unconstrained maximum, the likelihood ratio is bounded between zero and one.
- The likelihood-ratio test requires that the models be nested, i.e. the more complex model can be transformed into the simpler model by imposing constraints on the former's parameters.

Example 10 (T4Q8)**.**

Suppose that $X_1, \ldots, X_n$ denotes a random sample from the probability density function given by

$$f(x|\theta_1, \theta_2) = \begin{cases} \left(\frac{\theta_1 \theta_2^{\theta_1}}{x^{\theta_1+1}}\right), & x > \theta_2 \\ 0, & \text{otherwise.} \end{cases}$$

The following random sample of 10 has been observed:

143, 94, 163, 143, 140, 128, 112, 178, 124, 119

Determne the likelihood test statistc for testing $H_0 : \theta_2 = 89.1$ versus $H_1 : \theta_2 \neq 89.1$ with θ_1 unknown.

Example 11 (T4Q9)**.**

You fit a Pareto distribution to a sample of 300 claim amounts and use the likelihood ratio test to test the hypothesis that $\alpha = 1.9$ and $\theta = 6.9$. You are given:

- The maximum likelihood estimates are $\hat{\alpha} = 1.7$ and $\hat{\theta} = 6.7$.
- $\sum \ln(x_i + 6.9) = 639.61$
- $\sum \ln(x_i + 6.7) = 591.28$

Let Q be the value of the likelihood ratio test statistic and u be the degrees of freedom. Determine $Q - u$.

Example 12 (T4Q10)**.**

You fit a Gamma distribution with parameters $\alpha = 2$ and θ unknown to a sample of 5 claim amounts and use the likelihood ratio test to test the hypothesis $H_0 : \theta = 63.9$ vs $H_1 : \theta \neq 63.9$. You are given:

x_1	x_2	x_3	x_4	x_5
304.6	63.6	68.6	40.2	325.0

Determine the value of the likelihood ratio test statistic.

Example 13 (T4Q11)**.**

You fit a Weibull distribution with parameters $\tau = 2$ and θ unknown to a sample of 5 claim amounts and use the likelihood ratio test to test the hypothesis $H_0 : \theta = 10.9$ vs $H_0 : \theta \neq 10.9$. You are given:

x_1	x_2	x_3	x_4	x_5
36.2	4.1	4.6	1.9	39.3

Determine the value of the likelihood ratio test statistic.

4.3 Score Based Approaches

4.3.1 Akaike Information Criterion (AIC)

Suppose that we have a statistical model of some data. Let k be the number of estimated parameters in the model. Let $\hat{L}$ be the maximum value of the likelihood function for the model. Then the AIC value of the model is the following.

$$AIC = 2k - 2\ln(\hat{L})$$

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Thus, AIC rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting, which is desired because increasing the number of parameters in the model almost always improves the goodness of the fit.

Example 14 (T4Q12).

You are given a sample of 5 observations from $Pareto(\alpha, \theta = 1240)$ distribution:

$1,485.35 \quad 1,897.14 \quad 1,241.05 \quad 1,399.10 \quad 1,308.37.$

Determine the value of the Akaike Information Criterion (AIC).

Example 15 (T4Q13).

You are given a sample of 10 observations from the following distribution:

$$f(X) = \frac{1}{2\theta^3} x^2 e^{-x/\theta}, x > 0$$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
110.70	139.61	36.82	82.10	69.66	78.08	148.24	66.45	142.85	275.47

Determine the value of the Akaike Information Criterion (AIC).

Example 16 (T4Q14)**.**

A model is used for claim frequency. One of the explanatory variable is age. Age is a categorical variable. There are 12 age groups. Model I includes age as an explanatory variable and Model II does not. The models are otherwise identical. You are given:

Model	Likelihood Ratio Statistics	AIC
Model I	-20.2	139.8
Model II	-26.6	

Determine the AIC for Model II.

Example 17 (T4Q15)**.**

You fit various models for 21 loss observations using maximum likelihood. The fits maximizing the likelihood for a given number of parameters have the following loglikelihoods:

Number of parameters	Loglikelihood
1	-142.12
2	-141.68
3	-139.32
4	-138.12
5	-137.17

If AIC is the value of the Akaike Information Criterion, and K is the number parameters in the selected models. Find AIC+K.

Example 18 (T4Q16)**.**

You are given the following results for two models fit to the same data:

Model	AIC
$g(\pi) = \beta_0$	94.9
$g(\pi) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$	87.6

Calculate the likelihood ratio statistic to test $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

4.3.2 Bayesian Information Criterion (BIC)

Loglikelihood is proportional to the sample size, n. The likeihood ratio algorithm threshold will therefore be easier to meet as n grows. Thus, an alternative to likeihood ratio algorithm is the Bayesian Information Criterion (BIC). It is also called the Schwarz Bayesian criterion(SBC).
The BIC is formally defined as

$$BIC = k\ln(n) - 2\ln(\hat{L})$$

where

- $\hat{L}$ = the maximized value of the likelihood function of the model M, i.e. $\hat{L} = f(x|\hat{\theta},, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function;
- x = the observed data;
- n = the sample size;
- k = the number of parameters estimated by the model.

Model with the smallest BIC values will be selected.

Example 19 (T4Q17).

You fit a Pareto distribution with parametes α and $\theta = 50$ to a sample of 260 claim amounts. You are given:

- $\sum_{i=1}^{n} \ln(x_i + 50) = 1227.71$

Determine the value of the Bayesian Information Criterion (BIC).

Example 20 (T4Q18).

You fit a Gamma distribution to a sample of 10 claim amounts below.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
242.7	283.9	122.1	199.5	179.7	193.2	295.8	174.4	288.4	459.[illegible]

The maximum likelihood estimates are $\hat{\alpha} = 7$ and $\hat{\theta} = 34.84$. Determine the value of the Bayesian Information Criterion (BIC).

Example 21 (T4Q19)**.**

You are testing the addition of a new categorical variable into an existing model. You are given the following information:

- The change in the likelihood ratio statistic after adding the new variableis 56.
- The change in the AIC after adding the new variableis -36.
- The change in the BIC after adding the new variableis -35.

Calculate the number of observations in the model.

Example 22 (T4Q20)**.**

You fit various models for 20 loss observations using maximum likelihood. The fits maximizing the likelihood for a given number of parameters have the following loglikelihoods:

Number of parameters	Loglikelihood
1	-141.4
2	-140.84
3	-138.48
4	-137.48
5	-136.98

Using the Bayesian Information Criterion, how many parameters are in the selected models.